

Ein konzeptuelles Schema für Familiendaten

Pierau, Karl

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Pierau, K. (1991). Ein konzeptuelles Schema für Familiendaten. *Historical Social Research*, 16(1), 48-59. <https://doi.org/10.12759/hsr.16.1991.1.48-59>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

Ein konzeptuelles Schema für Familiendaten

*Karl Pierau**

Abstract: Large-scale data modelling is a very important and the actually creative step on the way to a data base. A common technique of data modelling is the entity-relationshipmodel. This technique yields, keeping to some formal rules, data structures which can directly be transformed into a logical data scheme of a network or a relational kind. In such a way the data base mirrors relationships close to reality and the data can be very compactly stored as well. The paper is demonstrating this technique using data from parish registers for family reconstruction.

Der Aufbau von Datenbanken wird gelegentlich, und nicht nur in den Sozialwissenschaften, mit einer gewissen Nachlässigkeit betrieben. Bereits vor zwanzig Jahren wurden von E.F. Codd Normalformen entwickelt, die als Instrument zur Begründung geeigneter Datenstrukturen zum Einmal-eins eines jeden Datenbankentwicklers gehören sollten. Die Normalformtheorie geht von gegebenen Relationen aus und liefert ggf. veränderte Relationen, die für die Anwendung in einem Datenbankmanagementsystem günstige Eigenschaften besitzen. Ob diese normalisierten Relationen tatsächlich ein geeignetes Abbild der zu untersuchenden realen Welt darstellen, hängt weniger vom Vorgang des Normalisierens, sondern in erster Linie von den Ausgangsrelationen ab. Wird in diesen der Untersuchungsgegenstand verzerrt dargestellt, z.B. könnten sie aus der Struktur des Quellenmaterials abgeleitet sein, so ist auch von den normalisierten Relationen keine adäquate Darstellung zu erwarten. Widerspiegelt die Darstellung im Computer aber nur ungenügend das Original, so wird man sich selbst zusätzliche Mühen bei der Auswertung bereiten. Das Problem besteht also darin, durch geeignete Entwurfsmethodiken, ausgehend von dem zu untersuchenden Ausschnitt der realen Welt, zu einem Abbild desselben zu kommen, das die wesentlichsten für die Aufgabenstellung relevanten Seiten wiedergibt. Dieses Abbild oder Modell soll sich möglichst einfach in Datenstrukturen, die ggf. normalisiert werden können, überführen lassen.

* Address all communications to: Karl Pierau, Borner Straße 11, O-1093 Berlin.

Ein dem Phasenkonzept der Softwareher Stellung analoges Herangehen wurde für den Entwurf und Aufbau von Datenbanken entwickelt. In einer ersten Phase werden die allgemeinsten Aufgaben formuliert, die Bedürfnisse und Sicht weisen der verschiedenen Nutzer ermittelt, Quellen erschlossen. Inhaltlich könnte man diesen Abschnitt als 'abstecken eines Ausschnitts der realen Welt' charakterisieren. Im zweiten Schritt muß dieser Ausschnitt durch geeignete Mittel formalisiert beschrieben werden. Diese Beschreibung wird als konzeptuelles Modell oder konzeptuelles Schema bezeichnet. Es wurde eine Vielzahl von Beschreibungsmitteln entwickelt. Ein sehr verbreitetes ist das Entity-Relationship-Modell (ERM), auch 'Gegenstands-Beziehungs-Modell' genannt, wobei die Übersetzung von Entity' in 'Gegenstand' nicht sehr glücklich getroffen scheint. Da die, z.B. bei Zehnder (1), gewählte Bezeichnung Entität' für den mit der Materie weniger vertrauten Leser noch unverständlicher ist, wird im folgenden 'Gegenstand' genutzt, wohl wissend, daß die solcherart klassifizierten Objekte nicht notwendig etwas Gegenständliches darstellen. Gegenstände im Sinne dieses Modells sind Objekte der realen Welt oder auch unserer Anschauung. Die Menge aller in einem gewissen Sinne gleichartigen Gegenstände bilden einen Gegenstandstyp. Das Modell hat seine Grundidee dem Systembegriff der Kybernetik entlehnt. Kurz gesagt, versteht man dort unter einem System eine Menge von Elementen, die in bestimmten Beziehungen zueinander stehen. Davon ausgehend wurden von verschiedenen Autoren spezielle Methoden entwickelt. An dieser Stelle wird das Modell von Wong und Katz, angewendet auf historische Familien, die unter demographischem und sozialhistorischem Gesichtspunkt untersucht werden sollen, vorgestellt. Es hat den Vorteil, daß das konzeptuelle Schema unter Beachtung weniger formaler Regeln direkt in normalisierte Relationen überführbar ist.

Die Daten wurden durch Familienrekonstitution aus Kirchenbüchern gewonnen und liegen als Stammfolgen vor. Ergänzendes Material sind Festschriften, Steuerlisten und Schullisten. Abbildung 1 zeigt ein Familienblatt aus den Stammfolgen.

Gegenstände und Beziehungen

Bestimmend für die Modellierung ist die Aufgabenstellung, im vorliegenden Fall mikroanalytische Untersuchungen zu demographischen und sozialhistorischen Problemen. Welches sind in diesem Sinne die grundlegenden Elemente oder Gegenstände? Männer und Frauen gehen Partnerschaften ein, von den Frauen werden Kinder geboren, die heranwachsen und in analoger Weise Familien gründen. Personen treten als Kind, Mann oder Frau auf. Da, über den ganzen Zeitraum gesehen, eine Person sowohl

Abb. 1 Familienblatt der Familie Allner
aus 'Stammfolgen in Nowawes-Neuendorf 1752-1874',
von B. Schwarz, Nowawes, 1937, Band 1

<i>Allner, Aug. Ferd. Ethel</i>		<i>Niether, Doreth. Luise</i>	
* <i>Beelitz</i> 1.1.1833		* <i>Kade</i> 16.12.1861	
+ <i>Now.</i> 15.2.1900		+ <i>Now.</i> 17.2.1900	
+ <i>Allner, Aug</i> <i>Mählstr., Potsd.</i>		+ <i>Niether, Joh. Andr. Fdch, Ethel,</i> <i>Häusler, Cade(Donthin)</i>	
<i>Wm.; Ndf Schmergow(61-64)</i> <i>Handlem.; Now.(81-00)</i>			
+* <i>Ndf.</i> 00 <i>Potsd.</i> 1860			
1	<i>Oskar Aug.</i>	10.8.61	
2	<i>Luise Emma Bertha</i>	15.12.63	15.10.64

Kind als auch Mann oder Frau sein kann, ist es sinnvoll zu generalisieren, 'Personen' als Gegenstände zu betrachten. In der Literatur wird in diesem Zusammenhang der Begriff 'Rolle' gebraucht. Ein Objekt aus der homogenen Menge Personen tritt in der Rolle 'Kind', 'Mann' oder 'Frau' auf. Formal gesehen kann ihr aus der Menge - 'männlich', 'weiblich', 'unbekannt' - ein Element zugeordnet werden. Das Geschlecht einer Person kann als die Beziehung zwischen dem Gegenstand Person und dieser Menge dargestellt werden.

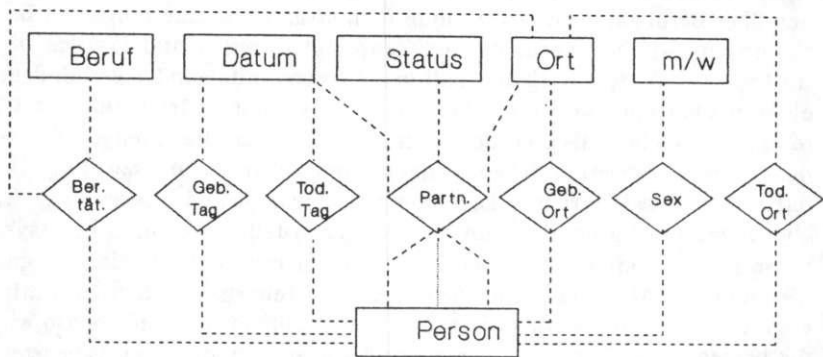
Welche anderen Mengen sind im Zusammenhang mit Personen' von Bedeutung? Die auf der Quellenbasis nachvollziehbaren Ereignisse traten zu bestimmten Zeitpunkten und an bestimmten Orten auf. D.h. die Menge der Personen steht mit einer gewissen Menge von Orten und einer Menge von Zeitpunkten (von Typ Datum) in Beziehung. Eine Person steht durch ihren Geburtstag und Sterbetag jeweils mit einem Datum in Beziehung. Analog dazu erkennt man durch Geburt und Tod eine Verbindung zwischen Personen und Orten. Personen sind durch ihre Berufstätigkeit zu einem bestimmten Zeitpunkt mit einem Ort, in dem sie ihren Beruf ausübten, verbunden. Eigentlich müßten wir hier Zeitperioden betrachten.

Da man aber nur zu bestimmten Zeitpunkten, zu den in den Kirchenbüchern registrierten Ereignissen Geburt, Heirat und Sterbefall, Informationen über Berufstätigkeit besitzt, muß man ersatzweise eine Folge von Beziehungen zwischen Personen und Zeitpunkten, Orten und Berufen betrachten. Es werden in diesem Fall vier Mengen miteinander verbunden: einerseits Personen und andererseits Beruf, Datum und Ort. Genauso hätte man die Geburt als eine Beziehung zwischen den drei Mengen Personen, Orte und Datum sehen können (und nicht Geburtstag und Geburtsort als zwei getrennte Beziehungen). Diese Sicht ist zwar möglich, aber nicht günstig im Sinne unserer Aufgabenstellung. Es muß beispielsweise für demographische Analysen das Alter ermittelt oder im Zusammenhang mit Migrationsuntersuchungen der Geburts- und Sterbeort miteinander verglichen werden. D.h. die beiden Zeitangaben sind ebenso wie die beiden Ortsangaben bei Geburt und Tod im Sinne der Aufgabenstellung viel enger miteinander verbunden als Geburtsdatum mit Geburtsort bzw. Sterbedatum mit Sterbeort. Daher ist die vorgeschlagene Modellierung durchaus naheliegend.

Nicht dargestellt wurden bisher verwandtschaftliche Beziehungen, also solche zwischen Personen. Dabei kann es nicht darum gehen, alle möglichen verwandtschaftlichen Beziehungen explizit darzustellen, sondern nur solche, die für die Aufgabenstellung wesentlich sind und die es gestatten, andere durch Verknüpfung herzuleiten. Besonders wichtig erscheinen partnerschaftliche Beziehungen zwischen Mann und Frau. Ausdrücklich werden hier nicht nur Ehen betrachtet, sondern auch außereheliche Beziehungen, die implizit in einem der Kirchenregister bzw. in den Stammbögen Erwähnung fanden. Mit Partnerschaften sind neben zwei Vertretern des Typs Person - dem Mann und der Frau - noch drei weitere Gegenstände verbunden: Orte als Heiratsort, Datum als Heiratsdatum (bei unehelichen Ort bzw. Zeitpunkt der Erwähnung der Beziehung in der Quelle) und ggf. als Trennungsdatum sowie ein Gegenstand 'Status der Ehe' mit den möglichen Wertausprägungen 'ehelich', 'unehelich' und 'unbekannt'.

Die andere wichtige verwandtschaftliche Beziehung ist das Eltern-Kind-Verhältnis als Beziehung zwischen zwei Personen, oder auch einer Partnerschaft und einer dritten Person. Wie später gezeigt wird, lassen sich aus diesen beiden verwandtschaftlichen Verhältnissen alle anderen wie Geschwister, Großeltern, Onkel und Tanten, Neffen und Nichten ableiten. Abbildung 2 zeigt die Beziehungen zwischen den Gegenständen graphisch. Die Kästchen repräsentieren Mengen, also die Menge der Personen, Berufe, Orte, Zeitpunkte und die Menge 'Typ der Partnerschaft'. Die Beziehungen werden durch Rauten dargestellt, wobei die gestrichelten Linien die dazugehörigen Mengen verbindet. Die Eltern-Kind-Beziehung als Verbindung zwischen einer Partnerschaft und einer Person wurde als gepunktete Linie eingetragen.

Abb. 2 Darstellung der Beziehungen zwischen den Mengen



Dabei bedeutet:

Ber. tät - Berufstätigkeit

Geb. Tag - Geburtstag

Tod. Tag - Sterbetag

Partn - Partnerschaft

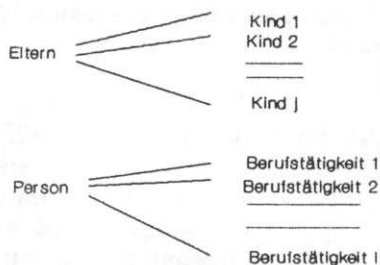
Geb. Ort - Geburtsort

Tod. Ort - Sterbeort

Das ERM von Wong und Katz

Wong und Katz schlagen vor, Gegenstände nach solchen zu differenzieren, die eine eigenständige Bedeutung im Sinne der Aufgabenstellung haben (strong entity), und solchen, die nur im Zusammenhang mit anderen Gegenständen von Bedeutung sind (weak entity). Die Differenzierung der Gegenstände legt auch eine Differenzierung der Beziehungen nahe. Assoziationen nennt man Beziehungen zwischen starken Gegenständen. Die zwischen starken und schwachen Gegenständen werden Attribute genannt. Als wesentlich und eigenständig kann man Personen sehen. Sie sind starke Gegenstände. Orte als Geburts- und Sterbeorte sind an Personen gebunden, haben nur im Zusammenhang mit diesen eine Bedeutung. Ebenso Zeitpunkte als Geburts- und Sterbedatum. Schwache Gegenstände werden auch, da sie im allgemeinen durch die Angabe eines Wertes dargestellt werden, z.B. Orte durch Angabe eines Ortsnamen, Zeitpunkte durch ein Datum, 'Werte' genannt. Starke Gegenstände nennt man dann kurz nur 'Gegenstände'.

Abb. 3 Bindungen



Aus einer anderen Sicht können Assoziationen durchaus eine gewisse Eigenständigkeit besitzen, z.B. Partnerschaften habe diese in unserer Untersuchung. Man billigt Assoziationen daher Attributierbarkeit zu, d.h. die attributive Verbindung zu einem Wert. Im betrachteten Falle bedeutet das, Partnerschaften werden über die Attribute Heiratsort, Heiratsdatum und Status jeweils Orte, Datumsangaben und ein Wert der über die Art der Partnerschaft Auskunft gibt ('ehelich', 'unehelich', 'unbekannt'), zugeordnet. Berufstätigkeit kann man an Personen gebunden sehen, ohne ihr eine eigenständige Rolle zuzuordnen. Man kann sie aber auch als wesentlich hervorheben, als eigenständig sehen auf Grund ihrer Bedeutung als Indikator für die soziale Einordnung der ganzen Familie. Es ist eine subjektive Entscheidung des Modellierers, wie er ihre Bedeutung sieht, welche Rolle er ihr zubilligt. Um das konzeptuelle Schema durch formale Regeln in eine Datenbankstruktur überführen zu können, muß an dieser Stelle eine Einschränkung gemacht werden. Jedem Gegenstand oder jeder Assoziation darf durch ein Attribut nur genau ein Wert zugeordnet werden. Da es mehrere Berufsangaben zu einer Person geben kann, muß die Berufstätigkeit in diesem Modell als starker Gegenstand betrachtet werden. Die Berufe, Berufsorte und die Zeitpunkte (der Erwähnung) der Berufsausübung sind mit ihm als schwache Gegenstände verbunden. Assoziationen, also Beziehungen zwischen starken Gegenständen, bei denen jedem Element der einen Menge genau ein Element aus einer anderen Menge zugeordnet ist, werden in diesem Modell 'Bindung' genannt. Z.B. hat jedes Kind genau ein Elternpaar (das u.U. nicht im Daten bestand enthalten ist). Eine weitere Bindung besteht zwischen Personen und Berufstätigkeit. Jede in einem der Kirchenbücher registrierte Berufstätigkeit ist genau mit einer Person verbunden. Aber eine Person kann natürlich mehrfach Angaben zu ihrer

Berufstätigkeit haben. Graphisch können die Bindungen wie in Abbildung 3 dargestellt werden.

Die Gegenstände auf der rechten Seite nennt man die gebundenen Gegenstände. In Tabelle 1 sind alle Gegenstände, Werte, Assoziationen und Attribute aufgelistet.

Tabelle 1

Gegenstand/ Assoziation	Attribut	Wert
Person	Geburtsdatum	Datum
	Sterbedatum	Datum
	Geburtsort	Ortsbezeichnung
	Sterbeort	Ortsbezeichnung
	Geschlecht	Zeichenkette
Berufstätigkeit	Beruf	Berufsbezeichnung
	Berufsort	Ortsbezeichnung
	Berufszeitpunkt	Datum
Partnerschaft	Heiratsort	Ortsbezeichnung
	Heiratsdatum	Datum
	Status	Zeichenkette

Tabelle 2

Assoziation/ Bindung	Gegenstände
Partnerschaft	Person
	Person
Eltern-Kind-Beziehung	Person
	Partnerschaft
Berufsausübung	Person
	Berufstätigkeit

Herleitung von Relationen

Das konzeptuelle Schema kann unabhängig vom beabsichtigten Einsatz eines Computers benutzt werden, um Einblicke in die Struktur des Untersuchungsobjektes zu gewinnen. Soll jedoch eine computergestützte Analyse durchgeführt werden, muß das noch von Datenverarbeitung unabhängige Schema in eine konkrete Datenstruktur auf einem konkreten Computersystem überführt werden. Es ist eine der Aufgabenstellung entsprechende Soft- und Hardware zu wählen. Datenbankaufgaben löst man zweckmäßiger Weise mit einem Datenbanksystem (DBS). Die heute vorzugsweise benutzten DBS basieren auf zwei logischen Grundmodellen. Das sind zum einen die netzwerkorientierten Datenbanksysteme, die die hierarchischen umfassen, und zum anderen die relationalen Systeme.

Durch die Übertragung des konzeptuellen Schemas auf eins der beiden logischen Datenmodelle (relational oder netzwerkorientiert) erhält man das logische Schema'. Hier soll das relationale Datenmodell benutzt werden, weil

- es einfacher handhabbar ist,
- der Untersuchungsgegenstand ohne wesentliche Verzerrungen darauf abbildbar ist, und weil
- einige relationale DBS vorliegen, die in Verbindung mit entsprechender Hardware eine kostengünstige Lösung der Aufgabe gestatten.

Der Begriff der 'Relation' wurde aus der Algebra übernommen. Der unbedarfte Leser möge sich darunter zur Veranschaulichung eine Tabelle vorstellen. Jede Zeile entspricht einem Element der Relation. In eine Spalte darf immer nur ein Wert aus dem vorher für diese Spalte definierten Wertebereich eingetragen werden. Z.B. könnten die Menge aller Dezimalzahlen kleiner 50000 mit zwei Nachkommastellen (monatliches Gehalt), oder die Menge aller Zeichenfolgen, die mit einem Großbuchstaben beginnen, dem eine endliche Menge von Kleinbuchstaben folgt (Nachnamen der Gehaltsempfänger), Wertebereiche sein.

Wenn bei der Herleitung des konzeptuellen Schemas gewisse von Wong und Katz geforderten Einschränkungen beachtet werden, kann man es unter Zugrundelegung formaler Regeln in ein logisches Schema überführen. Im folgenden werden die in Tabelle 1 und 2 dargestellten Gegenstände, Werte, Assoziationen, Bindungen und Attribute unter Anwendung von drei Regeln in Relationen übertragen, die dann ihrerseits sehr einfach in konkrete Datenbankstrukturen (physisches Schema) übertragbar sind.

Regel 1: Jeder Gegenstandstyp wird mit seinen Attributen und Bindungen zu einer Relation zusammengefaßt.

Regel 2: Jeder Gegenstandstyp wird mit einem Identifikationsattribut versehen, falls ein solches nicht schon vorhanden ist.

Regel 3: Jede Assoziation wird jeweils auf eine Relation abgebildet mit den Identitätsattributen der an ihr beteiligten Gegenständen und allen ihr zugeordneten Attributen.

Für den Gegenstandstyp Person wird eine Relation Person gebildet, die nach Regel 2 ein Identifikationsattribut Personennummer, die Attribute Geschlecht, Geburtsdatum, Geburtsort, Sterbedatum, Sterbeort und durch die Bindung an eine Partnerschaft, ihre Eltern, deren Identifikation erhält. Für den zweiten Gegenstand aus Tabelle 2, die Berufstätigkeit einer Person, bildet man die Relation Berufstätigkeit die das Identifikationsattribut Personennummer wegen ihrer Bindung zu Person erhält. Hinzu kommen die Attribute Beruf, Berufsort, Zeitpunkt der Erwähnung der Berufsausübung. Die Assoziation Partnerschaft erhält das Identifikationsattribut

Partnerschaftsnummer, Personennummer des Mannes, Personennummer der Frau und die Attribute Heiratsort, Heiratsdatum, Datum der Trennung der Partnerschaft, Status der Partnerschaft. Schematisiert dargestellt erhält man die folgenden Relationen:

PERSON (PERSONENNUMMER GEBURTSDATUM GEBURTSORT STERBEDATUM STERBEORT ELTERNNUMMER SEX)
BERUFSTÄTIGKEIT (PERSONENNUMMER BERUF BERUFSDATUM BERUFSORT)
PARTNERSCHAFT (PARTNERSCHAFTSNUMMER MANN FRAU STATUS HEIRATSDATUM HEIRATSORT TRENNUNGSDATUM)

Hier sind einige Anmerkungen notwendig. Orte wurden als Werte aufgefaßt. Selbstverständlich müssen gleiche Werte auch ein und denselben (schwachen) Gegenstand identifizieren. Damit kann als Wert des Ortes nicht der Ortsname fungieren, denn es gibt eine Reihe von Orten, die den gleichen Namen tragen. Es muß daher ein 'künstlicher' Name, ein Schlüssel, eingeführt werden, der jeden Ort eindeutig identifiziert. Der Nutzer möchte jedoch nicht einen verschlüsselten Ausdruck für eine Ortsbezeichnung vom Computer präsentiert bekommen, den er dann erst in irgendwelchen Tabellen nachschlagen muß. Daher müssen Orte neben ihrem eindeutigen Bezeichner, dem Schlüssel, noch das Attribut Ortsname bekommen. Ein Wert kann nicht attribuiert werden. Man ist also gezwungen Orte als Gegenstände aufzufassen, mit der attributiven Verbindung zum Wertetyp 'Ortsbezeichnung'. Bei der formalen Übertragung in Relationen sind die Berufs-, Heirats-, Geburts- und Sterbeorte dann Bindungen, die nach Regel 1 mit ihrem Identitätsattribut der Relation angefügt werden.

Bei den Berufsbezeichnungen kann einerseits nicht davon ausgegangen werden, daß hinter einem Beruf in einer längeren zeitlichen Entwicklung die gleichen oder zumindest ähnliche Tätigkeiten stehen. Zum anderen kann es für gleiche Tätigkeiten unterschiedliche Bezeichnungen oder zumindest Schreibweisen geben. Daher muß hier in analoger Weise wie bei Orten vorgegangen werden. Es empfiehlt sich für beide solche Schlüssel zu wählen, die Zusammenfassungen nach bestimmten Kriterien unterstützen. Das konzeptuelle Schema mußte nachgebessert werden, weil die zu modellierende reale Welt nicht scharf genug erfaßt worden war. Durch die zu grobe Zergliederung war das Modell einfacher und verständlicher. Man darf darin nicht einen Modellierungsfehler sehen; das beschriebene Phasenkonzept ist ein iterativer Prozeß. Treten zu irgendeiner Phase der Herleitung einer Datenbankstruktur Mängel auf, wird zu einem früheren Schritt gesprungen und die Prozedur läuft erneut ab. Personennamen wurden bisher außer acht gelassen. Als Identifikatoren für Personen sind sie ungeeignet, da sie im allgemeinen nicht eindeutig sind. Für viele historische Massenuntersuchungen sind sie nicht notwendig. Bei der Familien-

rekonstitution oder beim späteren Einbinden von zusätzlichen Informationen aus weiteren Quellen in die Datenbank, ist man jedoch gezwungen, sie zur Identifikation heranzuziehen. Im Sinne der dargelegten Methodik sind Namen attributiv mit Personen verbunden. Da auf sie relativ selten zugegriffen wird, sie jedoch die Relation 'Personen' stark aufblähen, empfiehlt es sich, Namen mit dem Identifikationsattribut für Personen in einer gesonderten Datei zu führen. Das heißt Namen als Gegenstände zu sehen, die durch eine Bindung an Personen gekoppelt sind. Das logische Schema in seiner endgültigen Form besteht aus sechs Relationen, die mit ihren Attributen im folgenden dargestellt werden.

Relation Personen	Relation Partnerschaften
Personennummer	Partnerschaftsnummer
Elternnummer	Mann
Geburtsdatum	Frau
Geburtsort (Schlüssel)	Heiratsdatum
Sterbedatum	Heiratsort (Schlüssel)
Sterbeort (Schlüssel)	Trennung
Geschlecht	Status
Relation Berufstätigkeit	Relation Orte
Personennummer	Ortsschlüssel
Berufsschlüssel	Ortsname
Berufsdatum	
Berufsort (Schlüssel)	
Relation Name	Relation Berufe
Personennummer	Berufsschlüssel
Name Berufsbezeichnung	

Prüfung der Relationen

Es ist zu prüfen, ob die Relationen ein geeignetes Modell der zu untersuchenden realen Welt darstellen. Sind sie es nicht, ist die Zergliederung erneut durchzuführen. Die Menge der Gegen**Standstypen** läßt sich in diesem Beispiel relativ leicht überschauen. Etwas schwieriger ist das schon bei der Rekonstruktion der verwandtschaftlichen Beziehungen. Abbildung 4 stellt die Beziehungen zwischen den Relationen graphisch dar.

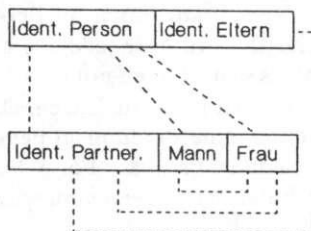
Zwei Arten von Beziehungen wurden in den Relationen explizit aufgenommen: Eltern-Kind-Beziehungen und partnerschaftliche Beziehungen. Es ist zu zeigen, daß aus diesen beiden durch Rekursion alle anderen rekonstruiert werden können. Wenn man zu einer Person alle Geschwister ermitteln kann, so sind über die Kette Kind-Mutter-Geschwister oder Kind-Vater-Geschwister alle Onkel und Tanten auffindbar. Nefen und Nichten findet man über Kind-Vater-Geschwister-Kind oder Kind-Mut-

Abb. 4 Personen und Partnerschaften mit ihren

Identitätsattributen

Personen

Partnerschaften



ter-Geschwister-Kind, die Großeltern sind die Eltern der Eltern. Es muß also für die Relationen nur gezeigt werden, wie man von einem Kind ausgehend die Eltern findet, von Eltern ausgehend alle Kinder, zu einer Person alle Geschwister und zu einer Person alle Partnerschaften ermitteln kann. Die Eltern einer Person findet man sehr einfach, indem man mit der Ident.-Nr. der Eltern aus der Relation PERSONEN in der Relation Partnerschaften den entsprechenden Satz herausucht. Geschwister sind alle die Personen mit der gleichen Elternnummer. Werden zu einer Frau alle ihre Partnerschaften gesucht, ermittelt man in der Relation PERSONEN die Ident.-Nr. der Frau, geht damit in die Relation PARTNERSCHAFTEN, Spalte Frau, und sucht dort alle die Sätze mit der Nr. der Frau. Damit hat man auch die Ident.-Nr.'n aller ihrer Partner. Mit diesen kann man aus der Relationen PERSONEN die Sätze der Männer herausuchen. Alle Kinder der Frau findet man, indem man aus der Relation PARTNERSCHAFTEN ihre Partnerschaftsnummern ermittelt und alle PERSONEN-Sätze herausucht, die diese Nummern als Elternnummer haben.

Damit ist gezeigt, daß die beiden genannten Beziehungen ausreichen, um alle anderen abzuleiten. Man sagt, sie sind hinreichend. Es läßt sich auch zeigen, daß beide notwendig sind, d.h. daß man keine von beiden weglassen kann ohne Informationsverlust hinnehmen zu müssen. Sie bilden ein hinreichendes und notwendiges System von Beziehungen. Bei korrekter Modellierung sind die Relationen in 3. Normalform. Es empfiehlt sich immer eine Prüfung daraufhin vorzunehmen um ggf. Fehler aufzudecken. Die Übertragung der Relationen in eine konkrete Datenbankstruktur für ein bestimmtes Datenbanksystem läßt sich relativ einfach realisieren. Darauf soll hier nicht weiter eingegangen werden.

Literaturhinweise:

- (1) Zehnder, C. A.; Informationssysteme und Datenbanken, Zürich 1985
- (2) Lockmann, P. C; J. W. Schmidt (Hrg.), Datenbank-Handbuch, Springer-Verlag, 1987